

Langley DAAC User Working Group Meeting

Executive Summary

March 26-27, 1998

Compiled by Lin Chambers

The meeting opened with a welcome and introductions, particularly of Dr. Cora Randall, our new co-Chair. Richard McGinnis (officially appointed DAAC manager in early April) then provided an overview of the issues to be covered during this meeting. These focus in particular on a number of changes which have taken place since our last meeting in the area of Federation, instrument launches, and the ECS system. The DAAC is now working with instrument teams (IT) in at least a half dozen different modes ranging from full production on DAAC (CERES on LaTIS), to routine production on DAAC (MOPITT at launch + 6 months), to no production on DAAC (SAGE archive and distribute only); with various other perturbations in support of field campaigns and joint projects.

CERES Review

Bruce Barkstrom then gave a review of the status of the CERES instrument and data processing to date; including an account of the Thanksgiving Day, 1997, launch. CERES data production is proceeding, but not without some unexpected problems (and some expected ones).

NASA Science Information Services (SIS) Study

Bruce Barkstrom also gave a report on the NASA SIS Study, in which he has been an active participant. That team has not yet made their report to headquarters, so comments are still welcome.

HDF

A number of topics relating to the Hierarchical Data Format (HDF) were discussed, beginning with a survey of the UWG members present. All had heard of HDF; three had actually ordered data products in HDF; and two had successfully read HDF data files. The report from the other side is "it's not that bad". A number of suggestions were made for improvements in the DAAC's documentation and instructions for HDF, to facilitate getting more scientists "over the hump". Bill North provided a report on a call for the sharing of Application Programming Interfaces (API's) which teams have developed to help work with their data. Only one team has contributed anything; and this is the team which made the suggestion. It was pointed out that CERES has created some of these, and Bill was requested to make another effort to get this type of tool shared.

A number of tools for looking at and reading HDF files were demonstrated. None of these were perfect; but each has been found to have some useful features. The JAVA HDF Viewer (JHV) by NCSA is useful to show a file's structure. It displays data as an image and can also show data values for a user selected portion of that image. `view_hdf` is an IDL-based tool developed for use within the CERES Project. It provides data plotting in a variety of plot formats as well as subsetting of data files. Since it has shown some promise as a more general use tool, discussions have been initiated about the possibility of making it available. EOSview is the EOSDIS-developed tool. It is useful for viewing HDF or HDF-EOS data in both tabular and image formats.

It was pointed out that while ideally HDF is a self-documenting file format, this only works well when the file writer makes use of these features. Examples were shown of HDF files that used default labelling, and ones which made full use of the labelling possibilities. The advantage of the latter for a user is quite obvious.



A number of resources for HDF were pointed out. A tools list at NCSA (including shareware and commercial HDF readers) is linked to from the DAAC. A listserv does exist for HDF questions.

Some actions were identified from this discussion; all imposed on the DAAC:

1. Add links from the DAAC to various resource sites (including page from last fall's HDF workshop) and to the e-mail listserv. Advertise it.
2. Generalize HDF instructions (i.e., not just for SUN workstation). Add an estimate of the time required for the installation and the disk space needed. Give some info on setting up paths or using SAs to help.
3. Link from data (at point of order) to minimal information (software, instructions, etc) needed to use it.
4. Provide a clearing house for questions and answers about HDF (archived and searchable). DAAC should provide discussion page monitor while ITs provide much of the information content.
5. Develop a bulletin-board type interface to make tools and experience available (not just restricted to HDF; but also other tools used to read data in various formats available on the DAAC).

Items for UWG Discussion

These agenda items were sprinkled throughout the two days in an attempt to keep us close to the schedule; but we did manage to address all of them.

Fostering Interdisciplinary scientists

This was a rather philosophical discussion without any real resolution about how to fulfill the mandate of EOSDIS, which is to facilitate interdisciplinary studies using the data collected. The example I cited was oceanography, where researchers are still using very simple equations for solar insolation; instead of some of the real surface radiation data available at the DAAC. Other examples abound, I am sure. Overall it was not felt that the DAAC could do much about this; except perhaps encouraging the User Services Working Group (USWG - an all DAAC coordination group) to make sure to emphasize radiation data at ocean conferences; and various other types of cross-field outreach. Effective use of the web and web-based search engines were also recommended, with a suggestion to study the idea of an EOSDIS-wide search engine which would lead a user to the appropriate DAAC's search and order tool. Also suggested was an article in EOS or BAMS (jointly by the DAAC and Science Teams) which would describe the DAAC's holdings and some possible uses. Brilliant ideas in this area will be warmly accepted.

More Science in the DAAC

This issue arose during the National Research Council (NRC) review of the DAAC, in which the recommendation was made to have a few PhD scientists at the DAAC to work with and find new uses for data in the archive. This point was hotly debated at the time, since the 100-odd person Atmospheric Science Division at Langley should fulfill this purpose. However, the recommendation appeared to stand. Further discussion among the UWG confirmed the DAAC's position that ASD fulfills this role for the Langley DAAC; but that there may be room to explore closer and new modes of collaboration in the future.

Co-Authorship for Data Providers

The UWG discussed this issue, which came up after a request from one data provider to the DAAC. All agreed that this should be handled as any other case: an acknowledgment (of the data provider and the DAAC) should be given or, preferably, a reference to the data source should be cited where available; co-authorship should be given only if the data provider made a significant contribution to the paper. Users should be encouraged to contact the data provider, especially for early release data. It was suggested that the DAAC maintain a list of potential reviewers for papers on its dataset holdings (perhaps from the instrument team members or websites). Also suggested was a pointer to a bulletin board-type system where the state of a given dataset can be discussed. This could include a list of what has been



done/checked/validated to date, and an assessment of the state of uncertainty of the data.

This issue does bring up once again the question of citable publications for the DAAC's holdings. The co-chairs were given an Action to work with Bob Seals on a review paper which would describe all the DAAC's holdings. (See also ["Interdisciplinary Science"](#) above.)

Link to UWG Member Page

There was a request from NASA management to make the list of members of the UWG more easily available to them. The group had no objections to making the web page visible to those who requested it, so that has now been done.

DAAC recruitment

We had another instalment in the discussion of how to help the DAAC recruit specialized people to work there. Suggestions this time included NCAR, AGU, AMS and AGU website, and the iwg-everyone mailing list. Brilliant ideas in this area will be warmly welcomed as well.

Issues at the DAAC

The Federation and future NASA support of Earth Science

The view from NASA HQ: more competition (i.e., lots of NASA Research Announcements); few if any unsolicited proposals funded; moving to PI-led missions (ESSP, etc). The NASA plan to implement NRC's recommendation on a self-governing federations includes a phased approach: 1) federate and recertify existing DAACs, 2) compete new work (12 each of Type II and III Earth Science Information Partners (ESIP) were recently selected). All these are to create a "working Prototype" federation. Into this prototype future work (beyond CHEM-1 mission) must fit. Note that the DAAC is involved in a support role in several of the winning ESIP teams. In this new approach, it is unclear what will happen with long term archival; and it was noted that long-term in this definition is more than three (3!) years. By congressional mandate, NOAA has this job. However, they will provide a lower level of service (basically, you must know the data is there) and it is likely the user will have to pay for data.

In sum, the goals of the current federation effort are unclear. Various participants seem to want different things. Additional meetings are ongoing to refine plans for the federation. One view has DAACs as a fall-back when the ESIPs are no longer interested (i.e., they do the boring/hard stuff).

Year 2000 and the DAAC

A brief discussion of this issue took place. The DAAC feels that it has no issues; but only time will tell. Since the meeting, higher levels of management have been tracking down information on this problem, so perhaps the UWG need not be concerned with it further.

Status of LaTIS

The Langley TRMM Information System (LaTIS) achieved a number of milestones since the last UWG meeting. Of particular note, it supported early CERES/ TRMM processing requirements, with some initial quick-look data processed the day after launch. Routine production of validated subsystems is now on-going, while SSI&T continues on remaining algorithm elements. Data products have been made available via the web to the CERES Science team. They will soon be publically available. The initial version of an automated product generation subsystem (using the Big Sur database schema) is now working. Further, the system includes an emergency component to support processing of EOS-AM data 6 months post-launch.

A question was raised about what ECS has and LaTIS does not, which allowed LaTIS to be developed on such a short



time scale. The following items are required in ECS; but not in LaITS:

- billing and accounting
- security (very good)
- one-stop shopping
- more services
- more automation
- flight operations
- communications

Production statistics are being kept to monitor elapsed time for production. With few exceptions, the production through CERES subsystem 1 has been with 24 to 36 hours of the end of the spacecraft day. A good fraction of this time is due to the fact that data are not received from the DAAC until several hours after the end of the spacecraft day.

Lessons learned from the LaTIS experience include:

- using state-of-the-art technology can be frustrating and time-consuming
- working storage is a precious resource
- allow frequent algorithm changes in early post-launch period
- manual operations can be error-prone (automation a priority)
- need early, consistent definition of file-naming convention
- test early and often to discover problems

LaTIS is being expanded with additional hardware and automation capability to handle CERES on EOS-AM-1. Input on desired types of media distribution hardware is solicited.

The CERES software team was also consulted for its views. Overall, they reported a good working relationship with the DAAC. The two groups worked together to substantially streamline the SSI&T procedures. Pitfalls they noted were different results from code run on two different - but supposedly identically configured - machines; trouble keeping consistent toolkit versions; and the versioning and coordination problem in general. They also pointed out that something that runs with test data will not necessarily run when given real data.

Improving & Evaluating Services

Status of Web Development

DAAC personnel did a demo of a new DAAC search and download tool - which ran in most any Internet browser - as an alternative to the current data access prototype. The idea was that some software would be downloaded to the user's machine, so it would be faster than doing maps and searches interactively over the web. They were wondering whether people's security concerns would limit use of such a thing. The UWG agreed that they should continue their work in developing it, but made some suggestions to help educate users to the potential risks and benefits of such a system.

Subsetting

A demo of the subsetting code developed for the CERES SSF product was given. Currently, this software will be customized for each type of dataset but it is hoped that a generalized approach can be developed. Currently the output is in ASCII; but studies are underway on whether to provide the output in HDF or some other format.

FIRE.ACE Activities

A brief report was given on the DAAC's activities in support of the FIRE Arctic Cloud Experiment (ACE). DAAC personnel will be in Alaska during the course of this field experiment to support the experimenters with web page development and updating. Ship reports, weather forecasts, aircraft mission summaries and flight logs are to be posted



on the Internet in near real-time. Satellite and model data will also be made available on this site. This represents a rather new mode of operation for the DAAC in relation to field experiments.

Status of ECS at LaRC

Current work is on ECS V.2. V.1 never arrived. Drop 3 was installed in February and does not support MISR. Drop 4 was to be installed in late March. It will support MISR; but not launch. It was delivered on schedule, so some functions had to be sacrificed. Several updates will be needed to support launch. The schedule for these is completely undefined. Mitigation strategies are being put in place to alleviate the problems that have occurred.

DAAC Status

Access Statistics

The number of DAAC customers is continuing to increase, nearly doubling from 1995 to 1997. They come from a growing number of countries (49 in 1997). Data volume delivered has more than quadrupled in 2 years, to nearly 4 TB in 1997, while projects and datasets supported have increased less than a factor of 3. The trend of 90-95% media delivery, with the rest electronic, has continued. A few users continue to order a large volume; while about 90% of users get less than 2 GB. It was pointed out again that many of these users might be well served by subsetting options, which could dramatically reduce the volume of data the DAAC needs to deliver. In terms of access method, the WWW interface is growing in use, but there are still a few customers who continue to use the GUI, ChUI, and V0 interoperable system.

New Datasets

From Sept. 1997 to March 1998, over 210 MB of data were added to the archive in 18 datasets. This does not include any CERES data, as that has not yet been made public. Currently there are 221 datasets on-line, amounting to 1.072 TB. Some of these datasets are currently available from the WWW site only; and not from the IMS or interoperable system. 206 of these datasets have guides, while only a few project, source, and sensor guides have been created so far. Priority was given to the dataset guides.

New Guide Format

A new abstract guide format has been developed as an abbreviated way to give users information on a dataset. A sample has been placed on-line for comment by UWG members as an [item for discussion](#).

DAAC Outreach

Trading Cards

A presentation was made on the first set of trading cards, on the topic of the Earth's Radiation Budget. The cards as well as the website were displayed. Card set are available by request for UWG members who were not present. Some suggestions for the website were also made. Additional card sets on aerosols and tropospheric chemistry are currently in the works.

Newsletter

Discussion on the target audience for the newsletter continued. It was suggested that users within the past 2 years (not just 1) be mailed the newsletter, as many felt that was a more reasonable time interval for the type of customers the DAAC has.



Other

There was discussion of the possibility of using summer faculty fellowships or graduate fellowships to bring some additional science presence to the DAAC. The possibility of closer interaction with ASD was also explored, including the possibility of internal fellowships. However, it was pointed out that ASD is only a small segment of the DAAC's customer base, so they need to maintain the right balance. We also discussed the possibility of using such fellowships to bring in computer science, information science, or library science people; rather than people in the strict atmospheric science discipline areas, as a way to bring new ideas to the archival and data distribution business. Bruce Barkstrom agreed to discuss this with the NCSA folks while he is there.

Miscellaneous

The group voted unanimously to have more data-related displays on the walls in the DAAC building, as time and resources permit. For example, displays prepared for meetings and conferences could be displayed in between trips rather than put into storage. Placement of additional sample images on the website was also favored.

Next Meeting

Topics for the next meeting should include:

- follow-up on subsetting status
- discussion on distribution of ancillary products along with data (proprietary issues, default settings, etc)

[^ Top of Page](#) | [UWG Home Page](#)

